

Intrinsic Signatures for Forensic Identification of SOHO Inkjet Printers

Zhi Li, Wanling Jiang, Daulet Kenzhebalin, Alexander Gokan, Jan Allebach
Purdue University, West Lafayette, IN, US

Abstract

Counterfeiting of currency globally remains a significant problem. And according to the authorities, a large portion of the fake currency is produced by Small Office Home Office inkjet printers. Therefore, a new inkjet printer forensics technology would be useful to identify the model of the source printer given a print sample. In our paper, we study the print patterns from 15 low-cost inkjet printers that are being sold on the market and examine test targets at the microscopic level. We design 4 printer intrinsic features including Dot Size, Dot Density, Average Distance to Nearest Dot, and Nearest Dot-Sector Density Function to characterize the behavior of inkjet printers. Furthermore, we extend our research and develop a machine learning based Printer Identification System. Unlike handcrafted features that have intuitive meaning to human viewers, an alternative set of intrinsic features are extracted from the Residual Neural Network, and based on the Neural Network features, a Support Vector Machine classifier is trained and is able to perform the printer model classification. Our evaluation shows that the proposed system produces robust and reliable results.

Introduction

In a 2006 report, The U.S. Secret Service estimated that 1 in 10,000 currency notes in circulation is a counterfeit¹. In Europe, Small Office Home Office (SOHO) inkjet printers now account for over 50% of the production of counterfeit currency notes. Authorities charged with tracking counterfeit currency to its source have a range of resources at their disposal. Even if these tools do not definitively identify the particular unit that was used to produce a counterfeit note, any information that they provide can prove to be a valuable aid to the investigation.

There are many possible approaches to forensic printer identification [1]. Some of these methods require labor intensive effort by a highly trained observer. Examples include inspection of prints under a microscope [2], chemical analysis of the inks used to print the suspect currency [3], and detection of spur marks from the gears used to advance the media through the printer [4]. Other methods are based on image analysis, including analysis of the structure of printed character glyphs [5, 6], analysis of page geometric distortion [7], analysis of halftone dot structure [8], and analysis of the memory contents of the suspect printing device, combined with analysis of the printed page [9].

In this paper, we focus on the development of intrinsic printer features for SOHO inkjet printers that are based on the analysis of the printer dot structure in highlight regions. In contrast to Ref. [8] above, which considers only laser electrophotographic printers that use periodic, clustered-dot halftoning patterns, here we consider the spatial arrangement and size of individual ink drops in dispersed-dot, aperiodic (stochastic) halftone patterns.

We also introduce the Printer Identification System (PIS) that can autonomously identify the source machine based on a given print sample. This system is made possible by the power of recent developments in Deep Neural Networks (DNNs). A number of effort has been made to combine DNN and halftoning. Moon *et al.* [10] characterized the Inkjet printer model using Deep Neural Networks. Ferreira *et al.* [11] used a combination of Neural Network classifiers and external classifiers to do laser printer classification on low resolution scanned images. Here, we use Residual Neural Network (RNN) [12] to classify high-resolution inkjet print captures.

Experiment Design and Sample Acquisition

The goal of this paper is to develop features that can serve as intrinsic signatures for SOHO inkjet printers. As mentioned above, inkjet printers use dispersed-dot aperiodic (stochastic) halftoning algorithms. We specifically choose to analyze the halftone dot patterns in highlight regions, since such regions most clearly illustrate the spatial pattern of dots (each dot corresponds to a single inkjet drop), and the size of these dots.

To support this research project, we purchased 16 inkjet printers ranging in price from \$30 USD to \$90 USD. This printer set consisted of 9 different models from 4 major SOHO inkjet printer manufacturers, namely HP, Cannon, Epson, and Brother. In addition, for three of these models, we purchased three units of each model in order to explore unit-to-unit variations within the same printer model. One of the printers was dead on arrival, so all of our experiments were based on 15 printers. As a means of identification, each of the units was assigned an alphabet ranging from A to H, among which Printers C, D, E are from the same manufacturer, and Printers F, G, and H are from another manufacturer.

Test Page Design

For our paper, we designed a test page consisting of constant-tone patches with gamma-corrected absorbance levels of 21%, 15%, 10%, 5%, and 0% for each of three colorants (CMY). Our analyses are entirely based on the constant-tone dot patterns. We printed the test page with each of the 15 target printers, and captured images of selected regions using a QEA PIAS-II camera (Resolution 7663.4 dpi with 3.2mm × 2.4mm field of view (1024 × 768 pixels). We found that when the cyan, magenta, and yellow level is 5%, the dots are dense enough so that their spatial relationship is prominent, yet dot coalescence is reasonably less frequent. For the simplicity of this paper, the 5% cyan, 5% magenta, and 5% yellow patches will be referred as *Triple 5* patches from now on.

Another two vital settings in our experiments are the printer driver settings and the media used. All the pages are printed in the best print quality mode, and the resolution is set to 600 dpi. We choose 600 dpi because this is the standard resolution for SOHO inkjet printers. Also more intrinsic features of the dot

¹For the full report, go to: <https://www.federalreserve.gov/boarddocs/rptcongress/counterfeit/counterfeit2006.pdf>

spatial relationships can be shown by controlling all the prints to have the same resolution. As for the media, since we target the currency counterfeiting issue, it is desired to study the printer behavior with similar media. Note that some countries have switched to polymer banknotes in an effort to combat counterfeiting and reduce costs. Printing on this special media is beyond our research area. Paper-based currency like the U.S. Dollar is what we try to simulate. According to the Bureau of Engraving and Printing², US paper currency is made up of 75% cotton and 25% linen. Therefore, in our experiment, we acquired two types of linen paper from Envelopes.com and Southworth. We also used Boise Multi-Use Copy paper in our experiment, as it represents plain paper that is daily used and most accessible among all the paper types.

Our second half of the experiments is for the classification. This test page consists of 40 repetition of a group of test targets. These test targets are 5% cyan, 5% magenta, 5% yellow, 5% black and *Triple 5*, respectively. These patches are designed to be about $3\text{mm} \times 2\text{mm}$ to fit in the field of view of the QEA PIAS-II. The bar on top of the patches is designed to help the user align the field of view of the camera with the lattice of printer-addressable points.

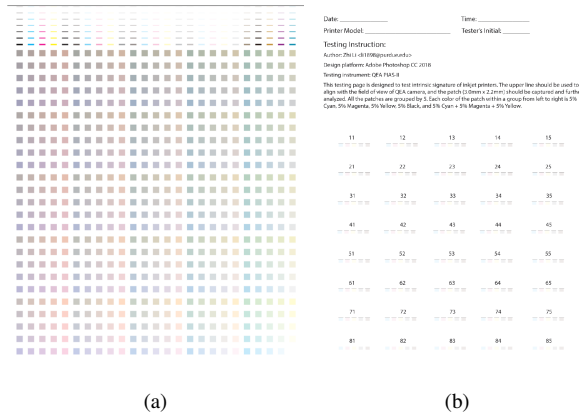


Figure 1: Designed test pages. (a) is the Phase I test page that is used to select a valid target; (b) is the phase II test page that contains same tone color patches replicated across the page.

Printer Characterization

We propose four different features to characterize the dot patterns from the captures shown in Fig. 2 Row 1: Dot Size, Dot Density, Average Distance to the Nearest Dot, and Nearest Dot-Sector Density Function. All our analyses are based on separate color channels, so we will do color separation first.

Colorant Separation

A colorant channel separation is desired to study the print-head characteristics individually. The following processes are used to obtain different channels:

Media pixel elimination. As shown in Fig. 2 Row 1, paper pixels are represented as white pixels in the captures, therefore we are able to identify the media/white pixels by examining a modified measure of color saturation S in CIELab color space

$$S = \frac{c^{*2}}{L^*}, \quad (1)$$

²For more information on U.S. currency and its paper and ink, see: <https://www.moneyfactory.gov/hmimpaperandink.html>

where c^{*2} is the square of the chroma value, and $c^{*2} = a^{*2} + b^{*2}$. We use c^{*2} instead of c^* in saturation calculation to balance the significance of the chroma and the lightness when the Lightness value is small. Since we only have yellow, magenta, and cyan ink, if a pixel is inked, it can only be one of, or a combination of these three primary colors (it is unlikely to have all three ink drops overlap together in such a light color, as shown in Fig. 2 Row 1). Therefore, as shown in Fig. 3, the saturation for inked pixels is higher, and other un-inked pixels have relatively low saturation value. Hence, we can eliminate the media pixels by thresholding S . Our experiment shows that when the threshold is 3, the separation performance is optimal.

Ink separation. To separate different ink colors, we measure the hue angle h_{ab} for every inked pixel.

$$h_{ab} = \arctan\left(\frac{b^*}{a^*}\right). \quad (2)$$

The quadrant is determined by the signs of the a^* and the b^* , and all the angles are calculated in radians. As shown in Fig. 4, the majority of the color pixels have the hue angle between $[-2, 2]$, where three peaks can be identified: (from left to right) cyan ink peaks around -2 , magenta ink peaks around -0.3 , and yellow peaks around 1.5 . Therefore we can separate the colored pixels by determining the closest peak. Some results are shown in Fig. 2. Although it is rare to see in these light color patches, ink overlapping and coalescence still can happen. We define the color pixels whose hue angles are between two adjacent peaks as overlapping pixels; and they count to both ink maps. Note that we are viewing the unwrapped histogram in Fig. 4. The real hue

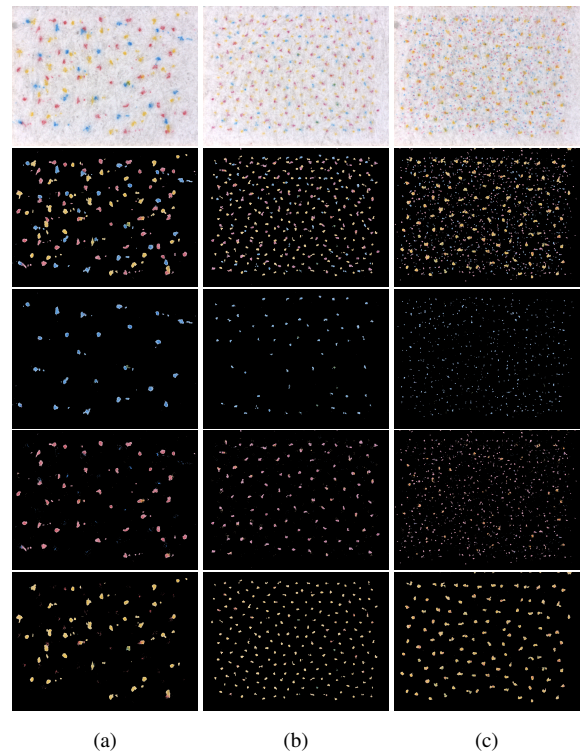


Figure 2: Sample separation results. The first row shows the original captures. The second row shows the images after eliminating the white media pixels. And the remaining three rows show the cyan, magenta and yellow channel, respectively. Column (a) shows the results for Printer B, Column (b) shows separation results for printer D, Column (c) shows the results for Printer H.

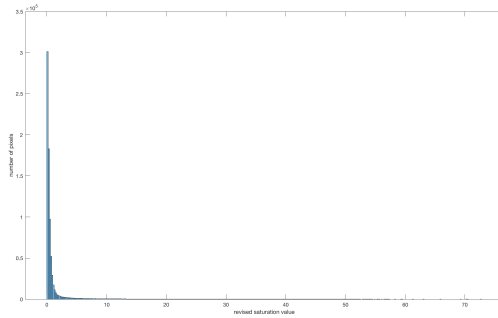


Figure 3: Distribution of pixels in capture images according to saturation. The saturation has a peak around 3 and a long tail as the saturation value increases.

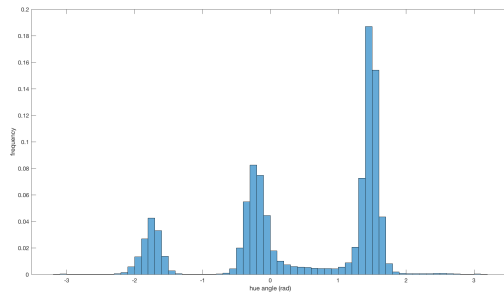


Figure 4: Distribution of the pixels in capture images according to the hue. The histogram has three peaks around -1.7, -0.3, and 1.5 (rad).

angle is distributed as a closed ring. Therefore, the cyan peak should also be adjacent to the yellow peak.

After the procedure illustrated above, we also would like to filter out noisy pixels. Firstly, insignificant clutters of pixels can be found in the captured images that are too small to be considered as an individual ink drop. Thus, we only consider dot-clusters that have more than 30 pixels. Secondly, pixels near a capture boundary should be eliminated from further study, as the full picture of their surroundings is unclear. Therefore, we only consider dots that are at least 50 pixels away from any of the boundaries. Then, we can perform connected component analysis to determine the different drops. Once we have a list of drops, data can be gathered about the drops such as their size, location, and compactness.

Dot Statistics

The print head and the ink used are critical aspects of the inkjet printer: print heads control the size of the ink drops and other behaviors of the jets, and the chemical nature of the ink determines the appearance of the prints and the ink spread on the paper, hence changing the dot shape. However, for a certain printer model, the print heads deployed and the ink selection are usually fixed, leading to consistent dot statistics as a unique feature shared within the same model. Therefore, we studied the dot statistics of the halftone image microscopic structure. And for each ink, we created the following features: Dot Area and Dot Density.

Dot Area

For Dot Area, we use the number of camera pixels to represent the size of the printed dot. Since we are using the same

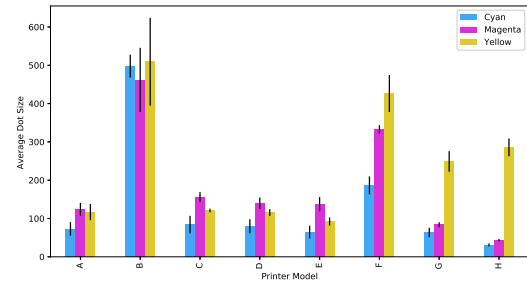


Figure 5: Average cyan, magenta, and yellow dot sizes comparison over all printer models.

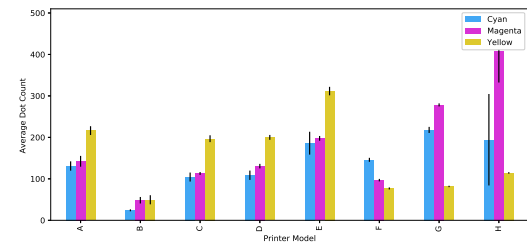


Figure 6: Average cyan, magenta, and yellow dot counts (number of dot pixel-clusters contained within the capture images) comparison over all printer models. The error bars show +/- 1 standard deviation.

magnification throughout all the captures, the number of pixels is proportional to actual dot size measured in mm^2 . The dot size comparison over all the models is shown in Fig. 5. It can be seen that the printers models from same manufacturer, for example, Printers C, D, and E, have similar dot sizes. Another three printers, Printers F, G, and H are also from the same manufacturer; and therefore their yellow dots are much bigger than are those for the other colorants. Note that Printer F is from a different series than Printer G and H; and the size of the dots from Printer F is quite different than the size of the dots from Printers G and H.

Dot Density

Another feature that we find to be distinctive is the dot density, or dot pixel-cluster count per capture. As shown in the Fig. 2, different printer models have not only very different overall dot densities (Row 2 in Fig. 2), but also different dot densities in the three different channels (Rows 3, 4, and 5 in Fig. 2). Hence, we evaluate all the captures and plot the bar chart on dot counts in Fig. 6. As seen in the figure, the dot densities of all the printers are very different: Printer B has overall the smallest number of dots across all 8 models and 3 colorants. Printers F, G, and H all have smaller numbers of yellow dots, but the portions of cyan, magenta and yellow dots are quite different; Printers A, C, D, and E have relatively similar proportions of dots across the three colorants with nearly the same densities for cyan and magenta. But the overall dot densities for Printer E are significantly larger than are those for Printers A, C, and D.

Spatial Distribution

Other than looking into the dot statistics, it is also very important to characterize the spatial dot distribution within each colorant. We propose two metrics to characterize the spatial distribution between the dots.

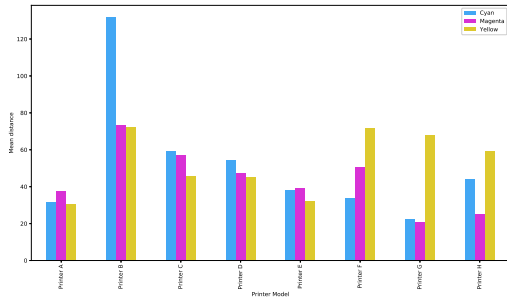


Figure 7: Comparison of Average Distance to the Nearest Dot (ADND) for the cyan, magenta, and yellow dot separations over all the printer models.

Average Distance to the Nearest Dot (ADND)

Average Distance to the Nearest Dot (ADND) is a measure of how close together the dots are. For each dot that is fully surrounded by its neighbors, we find out the distance between this dot and its nearest dot. Then, we are able to determine the ADND by average all the distances obtained in the capture image. As shown in Figure 7, we can see that Printers F, G, and H have a very distinct dot spatial arrangement compared to the others. The yellow dots are much sparser compared with other colorants. This is also consistent with the data shown in the Figs. 5 and 6, where we see that the yellow dots are much larger, and fewer in number for Printers F, G, and H, than for the other printers. Meanwhile Printers C, D, and E have similar spatial distributions, as they all from the same manufacturer.

Small ADND values also might also indicate a phenomenon that we call *dot pairing*. Dot pairing occurs when dots of the same colorant frequently occur in close, often nearly horizontal, proximity to each other. Note that unlike a major dot and its satellite dots, which is another distinct phenomenon that can be observed in high resolution captures, dot pairing is more consistent and the two dot sizes are relatively similar. It happens more frequently among the printers with smaller drop size. A set of examples are given in Fig. 8. As can be seen in the figure, cyan dots or magenta dots are paired together horizontally. The effect of dot pairing on the visual appearance of the print is not investigated in this study. But this phenomenon is very distinct and could be a good indicator of the printer model.

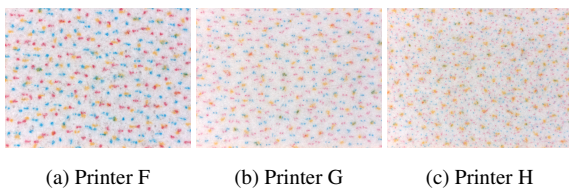


Figure 8: Examples of dot pairing. Printer F has the most frequent dot pairing phenomena, Printer G has less, and Printer H has the least.

Nearest Dot-Sector Density Function (ND-SDF)

The ADND metric draws a picture of how individual dots are spaced in the print: the biggest circle, centered at the centroid of the dot, which does not have any other dots inside of it. This concept of the dot placement measurement does not take into account the dot alignment or any other directional information. Therefore, we propose the second measure Nearest Dot-Sector

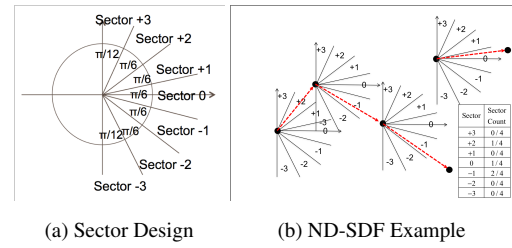


Figure 9: Nearest Dot-Sector Density Function (ND-SDF) concept

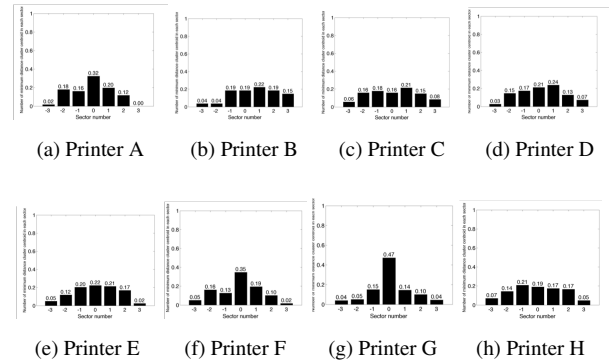


Figure 10: ND-SDF of cyan dots across all 8 printer models.

Density Function (ND-SDF). Inspired by the phenomenon of dot pairing, the ND-SDF metric is designed to capture the orientation of the dot and its nearest neighboring dot. The algorithm is described as below:

1. Set up 7 sectors as shown in Fig. 9.(a);
2. For color channel c , calculate S_c , the set of all the dots that are not near a boundary of the capture image field of view;
3. For each dot $d \in S_c$, search for the nearest dot in one of the sectors shown in Fig. 9.(a). If such a dot d' exists, do:
 - (a) Calculate the elevation or depression angle α between d and d' ;
 - (b) Register α to the corresponding sector i by adding 1 to count;
4. Normalize each individual sector value by dividing by the sum of all sector values.

Note that to avoid repetition, only dots in Quadrants I and IV are considered during the calculation. An example ND-SDF calculation is given in the Fig. 9.(b). The cyan ND-SDF histogram is shown in Fig. 10. We choose to show the cyan ND-SDF, as the cyan dots are more likely to pair together than other colorants. We can see that both Printers A and F have very high values in Bins -1, 0, and 1. This means that most of the nearest dot pairs accumulate around the same height. Our findings also correspond to what can be observed in Fig. 8.

Printer Identification System (PIS)

In the previous section, some hand-crafted features are engineered and designed to capture the characteristics of the microscopic structures from different prints. Although handcrafted features are intelligible to human examiners, there are some drawbacks:

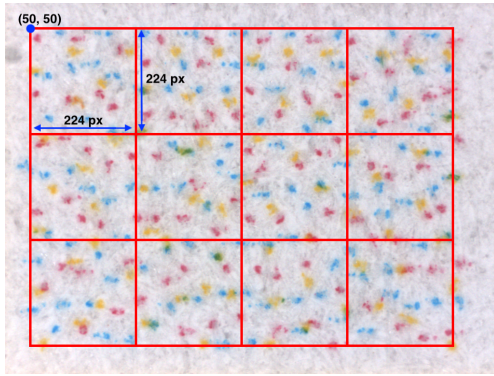


Figure 11: Example of retrieving 224×224 images from original capture.

1. It is rather tricky to model dot behaviors, as the inkjet imaging pipeline and inkjet marking engine technology is rather complex, and also stochastic rather than deterministic.
2. Handcrafted features cannot cover all the features seen and processed by the human viewers.
3. When future new printers join in the study, more features might need to be designed.

Therefore, a machine-learning based Printer Identification System (PIS) that can capture features autonomously is preferred for an anti-counterfeiting effort. Recent years have witnessed the rapid development of Deep Neural Networks (DNNs), and many DNN image object recognition applications have been used in our daily life. Hence, we aim to exploit the DNN's object recognition power to build a printer model classifier. We choose the Residual Network (RNN) [12], as it is one of the most popular and accurate networks in terms of image recognition³; and RNN also avoids some problems for network training. Here, we use ResNet50, a 50 layer version of the Network. Compared with ResNet50, other variations (101 layers or 152 layers) have deeper structures, which means that they are more prone to overfitting, while producing marginal accuracy gain.

First, steps should be taken to transform the data collected from the second phase experiment on 6 printers to data that can be used by the network. For each printer, we acquire more than 40 *Triple 5* patches using PIAS-II, and cut the each capture into 224×224 smaller images which is the acceptable input size of the network. As illustrated in the previous section, it is desirable to avoid the dot patterns around the image boundaries. Thus, we only collect 12 images from each original capture as shown in Fig. 11. Therefore, more than 3,500 image patches are collected as the training and testing dataset. Figure 12 presents some examples of the preprocessed input images.

Then, the pre-trained ResNet50 model is used to extract the features from the input images. Since all the weights in the model have been pre-trained using ImageNet [13], the model is well adapted to recognize both high level object details, as well as low level image features. However, original ResNet50 model outputs an label of an object in real life, for instance car, cat, dog, etc.; and this does not correspond to the desired labels in our task. Hence, we extract all the features from the second last layer and train another classifier, which is much simpler, lighter weight, and equally effective. The ResNet50 model produces

³ResNet50 won 1st place in the ILSVRC 2015 classification competition with top 5 error rate of 3.57%. For more information: <http://image-net.org/challenges/LSVRC/2015/index>

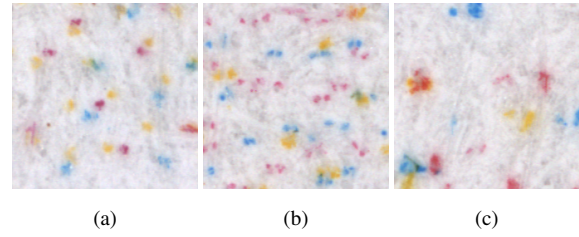


Figure 12: Sample ResNet50 input images. All the images are cropped to size 224×224 . Image (a) is from Printer D, image (b) is from Printer G, and image (c) is from Printer B.

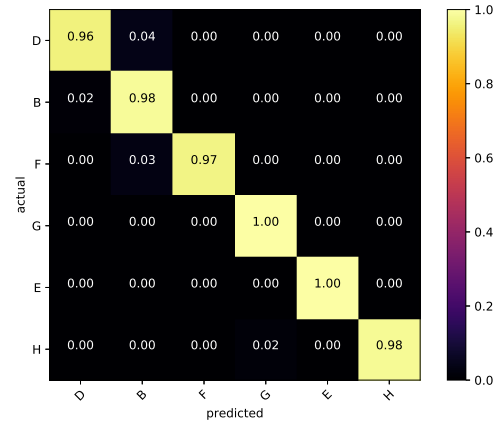


Figure 13: The confusion matrix of the classification results.

2,048 features for a single input image. But these features should not be directly used for the classification task for the following two reasons:

1. The dimension of the feature domain directly effects the computational complexity of the training process. With more features to take into account, more time is needed to train the classifier.
2. Using the full size feature sets will lead to overfitting. Considering the limited number of data points we have, a smaller feature set can help improve the generalization performance.

Thus, we use Principle Component Analysis (PCA) to reduce the feature space to 48 dimensions. We only use the first 48 features for training. Note that before conducting PCA, we shuffle and split the data into training and testing sets, and PCA model is fitted to the training set. Then, the same model is used to transform the testing set.

The final step is classification. We use Support Vector Machine (SVM) as the classifier model, and Radial Basis Function (RBF) as the feature space kernel. The confusion matrix of the testing results is shown in Fig. 13. The results here are promising: most of the prediction on the testing set is correct, especially for Printer G, H and E all the predictions are correct. Theoretically overfitting is avoided by using SVM classifier and PCA dimension reduction. However, we are working on more data to further test and validate the classifier.

Conclusion

In recent years, SOHO inkjet printers play an important role in modern office productivity, but they are also being used as one of the major counterfeiting tools around the world. In this paper, we investigate the intrinsic signatures of currently popular

SOHO inkjet printers. And we develop an identification system to predict the deployed printer by examining a printed document. We acquired 8 printer models from 4 major SOHO printer manufacturers. By examining the dot patterns at a microscopic level, we are able to design four intrinsic inkjet printer features that can capture both dot features and spacing features: Dot Size, Dot Density, Average Distance to Nearest Dot and Nearest Dot-Sector Density Function. We find that the printers made by the same manufacturers share more similarity than ones from other manufacturers. Our features are also able to capture phenomena like dot pairing. Finally, we use a Deep Neural Network to extract high dimensional intrinsic features from the collected prints, and fit a Principle Component Analysis model to reduce the feature set to 48 features. A SVM-based classifier is trained on the reduced feature set, and our testing results show the overall prediction accuracy is higher than 95%, and for some printer models the accuracy approaches 100%.

Acknowledgement

We want to thank Dr. Palghat Ramesh for suggesting the deep learning pipeline, and the Palo Alto Research Center (PARC) for providing computing resources. We also want to thank Zuoqian Xu for the inspiring discussions during our re-search activities.

References

- [1] P.-J. Chiang, N. Khanna, A. K. Mikkilineni, M. V. O. Segovia, J. P. Allebach, G. T. C. Chiu, and E. J. Delp, *Printer and Scanner Forensics: Models and Methods*, H. T. Sencar, S. Velastin, N. Nikolaidis, and S. Lian, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.
- [2] B. Hobbs and M. T. Kebir, "Non-destructive testing techniques for the forensic engineering investigation of reinforced concrete buildings," *Forensic Science International*, vol. 167, no. 2, pp. 167 – 172, 2007, selected Articles of the 4th European Academy of Forensic Science Conference (EAFS2006) June 13-16, 2006 Helsinki, Finland.
- [3] A. Rippert, "Overview on chemical analysis methods for soho inkjet inks," *SOHO InkJet Counterfeit Analysis Center Symposium*, 2018.
- [4] S. Georgescu, "Inkjet printers and their footprints on the paper/media spur marks analysis," *SOHO InkJet Counterfeit Analysis Center Symposium*, 2018.
- [5] E. J. D. Aravind K. Mikkilineni, Nitin Khanna, "Forensic printer detection using intrinsic signatures," *Proc.SPIE, Media Watermarking, Security, and Forensics III*, vol. 7880, pp. 7880 – 7880 – 11, 2011. [Online]. Available: <https://doi.org/10.1117/12.876742>
- [6] J. Aronoff and S. Simske, "Effect of scanner resolution and character selection on source printer identification," *Journal of Imaging Science and Technology*, vol. 55, no. 5, Sept. 2011. [Online]. Available: <http://dx.doi.org/10.2352/J.ImagingSci.Technol.2011.55.5.050602>
- [7] Y. Wu, X. Kong, X. You, and Y. Guo, "Printer forensics based on page document's geometric distortion," *Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP 2009)*, pp. 2909 – 12, 2009. [Online]. Available: <http://dx.doi.org/10.1109/ICIP.2009.5413420>
- [8] H. Wu, X. Kong, and S. Shang, "A printer forensics method using halftone dot arrangement model," *2015 3rd IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pp. 861 – 5, 2015. [Online]. Available: <http://dx.doi.org/10.1109/ChinaSIP.2015.7230527>

- [9] S. Fahd, M. Iqbal, M. Arif, and M. Javed, "Integrated model: Statistical features, memory analysis for scanner and printer forensics," *2016 4th International Symposium on Digital Forensics and Security (ISDFS)*, pp. 74 – 7, 2016. [Online]. Available: <http://dx.doi.org/10.1109/ISDFS.2016.7473521>
- [10] P. Moon, C. E. Kim, D. Kim, J. Moon, and I. Yun, "Ink-jet printing process modeling using neural networks," *Proceedings of the IEEE/CPMT International Electronics Manufacturing Technology (IEMT) Symposium*, 2008. [Online]. Available: <http://dx.doi.org/10.1109/IEMT.2008.5507800>
- [11] A. Ferreira, L. Bondi, L. Baroffio, P. Bestagini, J. Huang, J. A. Dos Santos, S. Tubaro, and A. Rocha, "Data-driven feature characterization techniques for laser printer attribution," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1860 – 1873, 2017. [Online]. Available: <http://dx.doi.org/10.1109/TIFS.2017.2692722>
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [13] R. Martinez, D. Smith, and H. Trevino, "ImageNet: a global distributed database for color image storage, and retrieval in medical imaging systems," *Proceedings. Fifth Annual IEEE Symposium on Computer-Based Medical Systems (Cat. No.92CH3117-9)*, pp. 710 – 19, 1992. [Online]. Available: <http://dx.doi.org/10.1109/CBMS.1992.244984>

Author Biography

Zhi Li is a fourth year doctoral student and teaching/research assistant in the School of Electrical and Computer Engineering at Purdue University, West Lafayette. He works with Prof. Jan Allebach and Poshmark primarily on fashion photography analysis, including fashion photography aesthetics and the autonomous garment color extraction system. He has also been involved multiple research projects such as fashion textural analysis, category/style classification based on CNN and more. Beyond academics, Zhi is an active member of the Purdue University Choir since 2014, and the Eta Kappa Nu (HKN) Beta Chapter since 2016. He is currently serving as the HKN volunteer director.